FROM BACH TO THE BEATLES: THE SIMULATION OF HUMAN TONAL EXPECTATION USING ECOLOGICALLY TRAINED PREDICTIVE MODELS

Carlos Cancino-Chacón, Maarten Grachten, and Kat Agres

Austrian Research Institute for Artificial Intelligence, Vienna
Department of Computational Perception, JKU Linz
Institute of High Performance Computing, A*STAR, Singapore

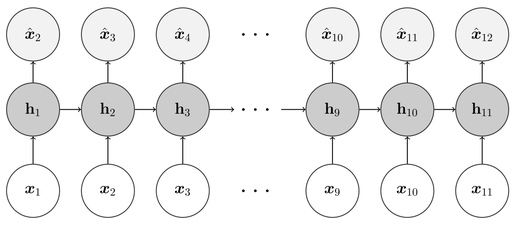
Introduction

- *Tonal expectation* plays an important role in music listening and understanding
- *Statistical models* have been shown to account for tonal perception, but often:
  - use reduced simplistic/artificial stimuli
  - no explicit account of acquisition of expectations through long term music exposure
  - use statical rather than sequential representations of musical contexts.

(NB: there are works that avoid some of these limitations, but none that avoid all at the same time)

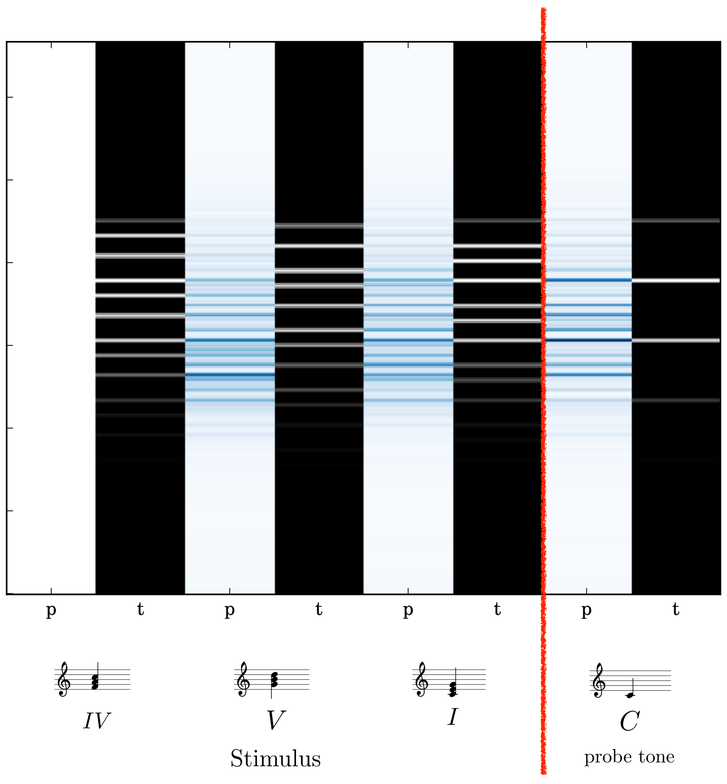
Method

- Train a RNN model to sequentially predict (beat-synchronous) constant-Q spectral slices on different corpora of real-world (commercially available) music:
  - 48 Preludes and Fugues of the Well Tempered Clavier (4 CDs, each pitch shifted 12 semitones)
  - 179 songs by The Beatles (13 CDs)



- Measure degree of tonal structure conveyed by model expectations, for models trained on different musical genres (by comparison with Krumhansl-Kessler (KK) key profiles [1])

Expectations of a model given a tonal context.
Blue on white columns: model predictions (p);
white on black columns: actually occurring CQT slice (t)



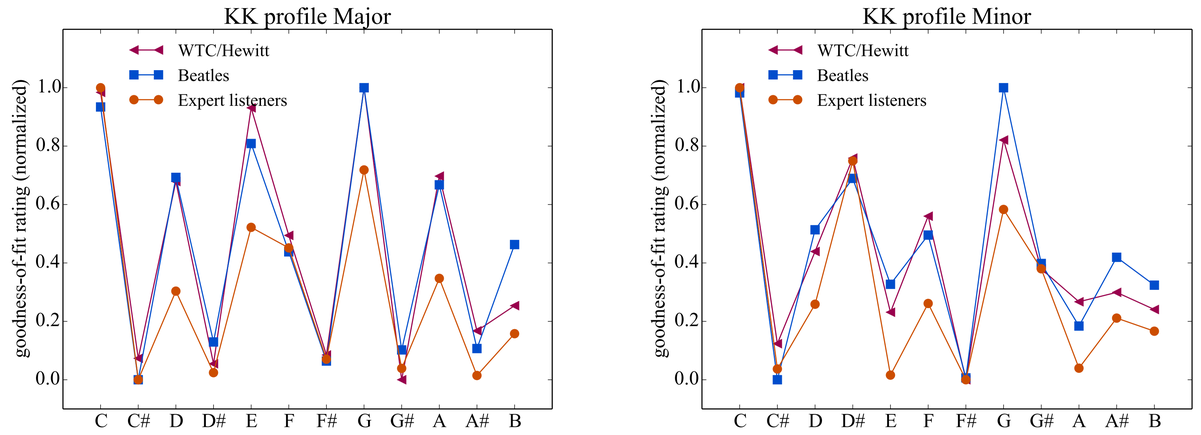
- **Investigate how the strength of tonal structure in expectations is related to the predictive accuracy of those expectations.**

Acknowledgements

This work has been partly funded by the European Research Council (ERC) under the EU's Horizon 2020 Framework Programme (ERC Grant Agreement No. 670035, project CON ESPRESSIONE). We thank Carol L. Krumhansl for providing the probe-tone data.

Results

Expectations of the models trained on WTC and Beatles datasets compared to average probe-tone ratings by expert listeners for major and minor contexts

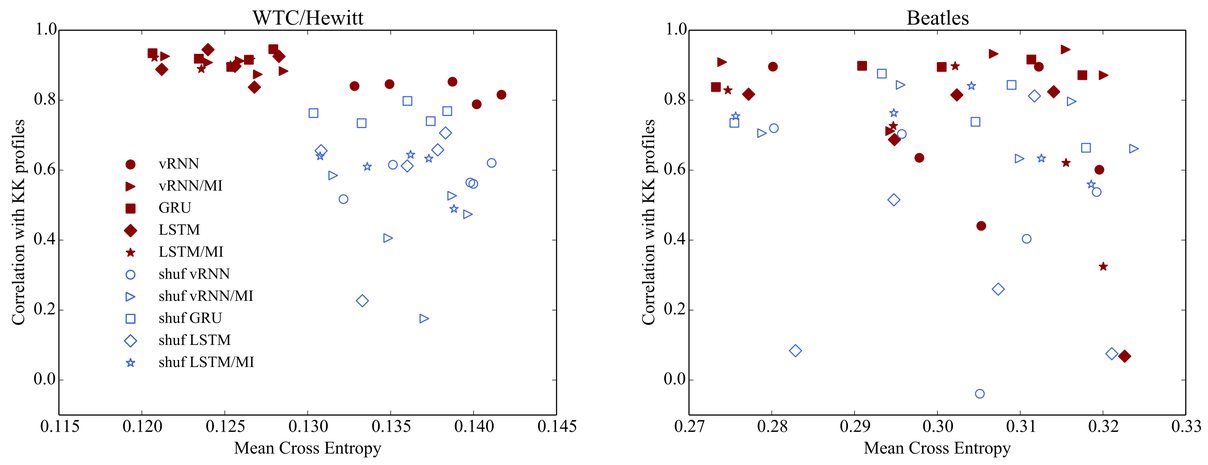


- The model achieves *high correlation with KK profiles* (as in previous studies, but with a more challenging setting, especially: model trained on data from all keys, not all transposed to single key)

Correlation between model predictions and KK profiles

	KK major	KK minor
WTC/Hewitt	0.915	0.940
The Beatles	0.900	0.885

Similarity of model expectations to human probe-tone ratings (Pearson's correlation coefficient) versus mean cross-entropy of the models over a test set



- Learning tonal structure is a necessary but not sufficient condition for accurate predictions.

Conclusions

- A sequential model trained on real-world musical stimuli to predict the future based on the current musical context (rather than find a subspace embedding of the current musical context, as in SOMs), develops tonally structured expectations similar to those obtained with simpler setups (see introductory points).
- The tonal structure in the expectations is an emergent effect of the training objective (to predict the future from the present musical context)
- Our experiments show that tonal perception is a necessary condition for accurate predictions, but not a sufficient condition: Accurate predictions likely develop through the interplay of tonal perception with other factors, such as perception of rhythm, and voice leading.

References

[1] C. L. Krumhansl and E. J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review*, 89(4):334–368, July 1982.

Contact: {carlos.cancino, maarten.grachten}@ofai.at, kat_agres@ihpc.a-star.edu.sg